Developing the Science of Privacy in support of the Art of Privacy

NSA Civil Liberties & Privacy Office

Feb 2015

Agenda

- The Problem: Civil Liberties & Privacy Assessments, Big Data, & Privacy Risk
- Back to Basics: What *is* Privacy?
- Developing the Science of Privacy
  - Challenge Questions for Research
  - Proposed Framework
- Next Steps

The Problem

- Big Data and Big Data Analytics challenge existing methodologies to evaluate privacy risk.
  - Every newly introduced data set can upend prior assumptions of privacy risk.
  - Every new analytic or combination of analytics in a workflow can upend prior assumptions of privacy risk

How can one build a scalable and manageable CLP assessment process in the Era of Big Data?

- "Privacy" as a concept is amorphous, legalistic, and deeply personal
  - The Right to be Forgotten? The Right to Hide? The Right to Conceal? No Right at All?
  - Is Meta-Data public or private (e.g. Smith v. Maryland)?
  - Is my data *my data* or is it the *intellectual property* of the service provider?

If there is no baseline of what privacy actually is, how can personal information be identified and effectively protected?

Our Goal

Develop a *practicable* approach to implement privacy protections.

- **Establish a *common lexicon* for *data* and *use*.**
- ***Assumption:* Privacy is a *Data-Driven* and *Use-Driven calculation*.**
- ***Assumption:* Built upon existing compliance and security framework**
- ***Assumption:* Privacy is the means by which one protects Civil Liberty (aka Individual Liberty aka Free Will aka Self-Determination) upon which the United States is founded.**

Data + Use ➔ Identify and Quantify Privacy Risk

Back to Basics

- What is *personal information*?
- What is *use* and what does it mean to *use* personal information?
- What really is *privacy risk*?
- How to handle context?

Developing the Science of Privacy

- The Science of Privacy is a ***principled*** and ***methodological*** approach to evaluate privacy risk, **using the scientific method**.
- Create a ***framework*** to underpin a ***Privacy Decision Support Tool***.
- Identify & Understand Privacy Risks

Research Challenge Questions

1. What are the actual privacy risks that need to be considered?

2. Can a mathematical method be developed to evaluate privacy risk based on the type of personal information present and the type of use(s) of that personal information?

3. How can an *Accountable Privacy Framework* be created for Big Data, building upon an existing compliance and security framework, that evaluates privacy risk based on *the type of personal information* and *type of use(s) applied*?

4. How can we apply current advances in privacy engineering? (e.g. Digital Rights Management, Differential Privacy, Homomorphic Encryption, Secure Multi-Party Computation)

Framework for the Science of Privacy

- Have begun initial investigations into potential ways to quantify risk of privacy in big data.
- Following is a proposed methodology.
- Focus is on *practicality* and *intuitiveness*.
- This initial methodology is a *work in progress* and readily elucidates opportunities, gaps, and challenges towards developing a framework for the Science of Privacy.

Framework: Personal Information Taxonomy

- Avoided definition of privacy
- Focusing instead on a broad definition of *Personal Information* such that it includes:

Any tangible information that can be used to identify an aspect of a person. (To include specific facts such as a name or address as well as patterns of behavior.)

- Attempting to apply IC "Identity Data Types" Taxonomy:
  - **Biometric**: Measurable, physical characteristics of an individual. (e.g., fingerprint, blood type, gait, gender).
  - **Biographic**: Attestable facts about an individual's life. (e.g., name, address, religion).
  - **Contextual**: Identity data from individual's transactions. (e.g., financial, commercial transactions, personal patterns).
- Assigning (*very!*) tentative relative privacy risk for each category.

Creating a Common Lexicon

DATA

- **Type**
  - Personal Information
    - Biographic
    - Biometric
    - Contextual
- **Bulk or Targeted**
  - *Targeted*: Known, identified threat
  - *Bulk*: Known targets, unknown targets, and innocent individuals intermixed
  - *Gradation as well!*

USE

- **Purpose**
  - Counter-Terrorism
  - Counter-Proliferation
  - Counter-Intelligence & Intents of Foreign Governments
  - Cybersecurity
  - Transnational Criminal Threats
  - Threats to Military & Allies
- **Analytical Activities**
  - Discovery
  - Targeted Collection
- **Technological Function**
  - Correlate
  - Filter
  - Format
  - Disseminate
  - Collect/Acquire

Identifying Possible Privacy Risks

1. Customer Information Needs
2. Analytic Strategy
3. Collection Strategy
4. Exploitation & Production
5. Dissemination
6. Feedback

Framework: Use Taxonomy

- Have identified a handful of hierarchies of "use".
  - *Purpose & Analytical Activities* – more *subjective*, per business needs.
  - *Technological Functions* – more *objective*, per analytical processes.
- Focusing initially on *Technological Functions*:
  - Analytics decompose into atomic technological functions (e.g., filter, correlate, etc.).
  - Composite analytic workflows can be constructed from individual analytics, each consisting of atomic technological functions.
- Assigning (*very!*) tentative privacy risk of use
  - e.g., Tech. Function: Correlation ➔ Raises Privacy Risk
  - e.g., Tech. Function: Filter ➔ Lowers Privacy Risk

Framework: Conceptual Diagram of Data & Analytic Use